Available online at www.sciencedirect.com
SCIENCE DIRECT®
Microbes and Infection 5 (2003) 1082–1095
Microbes and Infection
www.elsevier.com/locate/micinf

Original article

The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization

Sabine Daugelat [a,*], Jane Kowall [a], Jens Mattow [a], Dirk Bumann [b], Ralf Winter [c], Robert Hurwitz [c], Stefan H.E. Kaufmann [a]

[a] *Department of Immunology, Max Planck Institute for Infection Biology, Schumannstrasse 21–22, 10117 Berlin, Germany*
[b] *Department of Molecular Biology, Max Planck Institute for Infection Biology, Schumannstrasse 21–22, 10117 Berlin, Germany*
[c] *Biochemistry Core Facility, Max Planck Institute for Infection Biology, Schumannstrasse 21–22, 10117 Berlin, Germany*

Received 17 April 2003; accepted 16 July 2003

Abstract

A 9.5-kb section of DNA called region of deletion 1 (RD1) is present in virulent *Mycobacterium tuberculosis* strains but is deleted in all attenuated *Mycobacterium bovis* BCG vaccine strains. This region codes for at least nine genes. Some or all RD1 gene products may be involved in virulence and pathogenesis, and at least two, ESAT-6 and CFP-10, represent potent T- and B-cell antigens. In order to produce the entire set of RD1 proteins with their natural posttranslational modifications, a robust expression system for *M. tuberculosis* proteins in the fast-growing saprophytic strain *Mycobacterium smegmatis* was developed. Our system employs the inducible acetamidase promoter and allows translational fusion of recombinant *M. tuberculosis* proteins with polyhistidine or influenza hemagglutinin epitope tags for affinity purification. Using eGFP as reporter gene, we showed that the acetamidase promoter is tightly regulated in *M. smegmatis* and that this promoter is much stronger than the widely used constitutive *groEL2* promoter. We then cloned 11 open reading frames (ORFs) found within RD1 and successfully expressed and purified the respective proteins. Sera from tuberculosis patients and *M. tuberculosis*-infected mice reacted with 10 purified RD1 proteins, thus demonstrating that Rv3871, Rv3872, Rv3873, CFP-10, ESAT-6, Rv3876, Rv3878, Rv3879c and ORF-14 are expressed in vivo. Finally, glycosylation of the RD1 proteins was analyzed. We present preliminary evidence that the PPE protein Rv3873 is glycosylated at its C terminus, thus highlighting the ability of *M. smegmatis* to produce *M. tuberculosis* proteins bearing posttranslational modifications.

Keywords: *Mycobacterium tuberculosis*; Protein; Virulence; Glycosylation

1. Introduction

When the genome of *Mycobacterium tuberculosis* H37Rv was sequenced, 3924 open reading frames (ORFs) were annotated [1]. Experimental evidence and re-annotation has increased this number to 3995 protein-coding ORFs [2]. A function could be attributed to 58% of the ORFs of *M. tuberculosis*, 27% had no homology to ORFs outside the genus *Mycobacterium* (designated "conserved hypotheticals"), and 15% had no homology to any known genes [3]. Even though several other mycobacterial genomes have been sequenced since then, our understanding of the biology of *M. tuberculosis* is still limited. Some of the problems impeding mycobacterial research are: (i) the slow-growing, pathogenic mycobacteria are tedious to study, (ii) generating targeted gene deletions was almost impossible until recently and (iii) recombinant mycobacterial proteins have been notoriously difficult to produce in *Escherichia coli*. Mycobacteria have a high GC content (65.6% in *M. tuberculosis*), unique codon preferences and the encoded proteins tend to be rich in glycine, alanine, proline and arginine, all of which can cause problems for overexpression in *E. coli*. By 1992, only 52 mycobacterial proteins had been studied, mainly as T- or B-cell antigens, while their biological functions often remained elusive [4]. The majority of these proteins had been identified by N-terminal sequencing, but many had neither been cloned nor expressed in recombinant form. This became more fea-

* Corresponding author. Present address: Novartis Institute for Tropical Diseases Pte Ltd., 1 Science Park Road, #04–14 The Capricorn Singapore Science Park II, Singapore 117528, Singapore. Tel.: +65-6722-2925; fax: +65-6722-2910.
E-mail address: sabine.daugelat@group.novartis.com (S. Daugelat).

doi:10.1016/S1286-4579(03)00205-3

sible with the completion of the *M. tuberculosis* genome sequence, but currently only 24 recombinant *M. tuberculosis* proteins are available through the "TB Research Materials and Vaccine Contract" of the NIH (http://www.cvmbs.colostate.edu/microbiology/tb/top.htm) and it is safe to assume that no more than a few hundred full-length genes have been cloned to date and even fewer gene products have been expressed in recombinant form. As a consequence of the difficulties in expressing recombinant *M. tuberculosis* proteins, few monoclonal antibodies have been generated and are being used to study the *M. tuberculosis* proteome [5].

The 9455-bp region of deletion 1 (RD1) of *M. tuberculosis* is absent in all attenuated *Mycobacterium bovis* BCG vaccine strains due to a natural deletion event [6,7]. Therefore, it has long been speculated and recently been confirmed that the RD1 region contributes to bacterial virulence [8,9]. The RD1 gene products are interesting for various reasons: they represent potential virulence factors, vaccine candidates and may also be useful for diagnostic purposes. Little is known about the RD1 proteins; even their exact number is uncertain. Nine RD1 genes have been annotated in the genome project [1], but careful analysis has revealed the presence of up to 14 ORFs within this region [10]. One of these additional ORFs, designated *orf14*, was included in the present study because of the potential diagnostic usefulness of its gene product, ORF-14. Rv3872 and Rv3873 belong to the PE/PPE family of proteins. Members of this protein family are thought to be associated with the cell surface of *M. tuberculosis* [11,12], and some may bind fibronectin [13]. It has been postulated that PE/PPE proteins contribute to antigenic variation [1] or may represent spider silk-like structural proteins [14], but no function has been attributed as yet. The best-studied RD1 proteins are CFP-10 [15–17] and ESAT-6 [18,19], both of which elicit strong T- and B-cell responses in experimental animals and humans [16,18,20,21]. Despite the lack of classical signal sequences, both proteins are found in culture filtrates and might be exported by unknown mechanisms [22]. The *lhp* and *esat-6* genes (encoding CFP-10 and ESAT-6, respectively) are organized in an operon [17], their gene products are co-transcribed and form a tight 1:1 protein complex [23], but their biological functions remain unknown. The *lhp* and *esat-6* genes as well as other RD1 genes are part of a cluster called the "ESAT-6 superfamily", which is found in multiple copies in several mycobacterial genomes [24] and may contain a novel secretion apparatus [25]. In this context, it is noteworthy that Rv3871 has 2 ATP/GTP-binding motifs and is similar to FtsK/SpoEIII proteins. Rv3877 has 11 predicted transmembrane domains [26] and may be a transporter. So far, no function could be attributed to Rv3876, Rv3878, Rv3879c and ORF-14.

In order to study the RD1 gene products, we have created a family of vectors that allow expression of *M. tuberculosis* proteins in *Mycobacterium smegmatis*, a fast-growing saprophytic strain, by utilizing the inducible acetamidase promoter and various epitope tags to facilitate affinity chromatography. Using this system, we expressed 11 *M. tuberculosis* proteins (the nine annotated RD1 proteins, ORF-14 and the C-terminal domain of Rv3873, designated Rv3873T). Ten of these recombinant *M. tuberculosis* proteins were purified by chromatography and their identity was confirmed by mass spectrometry. We demonstrate that all 10 RD1 proteins reacted with sera from tuberculosis patients and *M. tuberculosis*-infected mice, thus indicating that they are expressed in vivo. Preliminary evidence is presented that the PPE protein Rv3873 is glycosylated at its C terminus, thus emphasizing the advantage of expressing recombinant *M. tuberculosis* proteins in *M. smegmatis* rather than in *E. coli* expression systems.

2. Materials and methods

2.1. Strains and media

E. coli was transformed by conventional heat shock transformation or electroporation and plated on LB agar (Invitrogen, Paisley, UK) containing appropriate antibiotics (hygromycin B [Roche, Mannheim, Germany] at 150 μg/ml, kanamycin [Sigma Aldrich, St. Louis, MO, USA] at 35 μg/ml or ampicillin [ICN, Aurora, OH, USA] at 100 μg/ml). For blue–white selection, 1 mM IPTG (Gerbu, Gaiberg, Germany) and 75 μg/ml X-Gal (Roth, Karlsruhe, Germany) were added. All cloning was done in *E. coli* DH5α (Invitrogen) unless otherwise stated. Liquid *E. coli* cultures were grown in LB medium (Invitrogen) containing appropriate antibiotics (hygromycin B at 150 μg/ml, kanamycin at 50 μg/ml or ampicillin at 100 μg/ml). *M. smegmatis* mc^2155 was electroporated as described [27] and plated on Middlebrook 7H10 agar (BD, Franklin Lakes, NJ, USA) supplemented with 10% albumin–dextrose saline (ADS: 0.81% NaCl, 5% BSA fraction V (Serva, Heidelberg, Germany), 2% glucose), 0.5% glycerol, 0.05% Tween-80 (Sigma) and either hygromycin B (50 μg/ml) or kanamycin (25 μg/ml). Liquid *M. smegmatis* cultures were grown in Middlebrook 7H9 medium (BD) supplemented with 10% ADS, 0.05% Tween-80 (Sigma) and 0.2% glycerol (further referred to as 7H9 complete medium), containing either hygromycin B (50 μg/ml) or kanamycin (25 μg/ml). For expression of recombinant proteins, *M. smegmatis* was grown in Middlebrook 7H9 medium (BD) without ADS but supplemented with 0.05% Tween-80 (Sigma), 0.2% glucose (Sigma) and 0.2% glycerol (further referred to as 7H9 expression medium).

2.2. Construction of pSD24

A promoterless episomal *E. coli*–mycobacteria shuttle plasmid, pMV206-Hygro (kind gift of W.R. Jacobs Jr.), was used as parent vector. The acetamidase promoter was amplified from *M. smegmatis* mc^2155 chromosomal DNA. A pair of primers named "ace-prom-Start" 5'-TTTAAAgaagtgac

gcggtctcaagcgtc-3' and "ace-prom-End" 5'-TTTAAAaac tacctcgggcatgtggactc-3' (capital letters indicating the *Dra*I restriction sites) was designed to amplify a 2.6-kb DNA fragment which included all putative regulators and the first 18 bp of the acetamidase gene *amiE*, according to published sequence information [28,29]. A PCR product was generated with *Pfu* Turbo DNA polymerase (Stratagene, Cedar Creek, TX, USA) and cloned into pPCR-Script SK$^+$ Amp, using the respective cloning kit (Stratagene). Sequence identity was verified by sequencing the first 500 bp of each junction. The 2.6-kb acetamidase promoter fragment was cut out by *Dra*I digest and blunt-end cloned into the *Eco*RV site of pMV206-Hygro to obtain pSD24. Restriction enzymes were purchased from New England Biolabs (Beverly, MA, USA); modifying enzymes were purchased from Roche. The resulting 6.8-kb plasmid, pSD24, allows construction of translational fusions regulated by the inducible acetamidase promoter (Fig. 1).

2.3. Construction of epitope-tagged expression plasmids pSD21, pSD22, pSD26, pSD29 and pSD31

The double-stranded oligonucleotide "Myco-C-His" was generated by hybridizing "myco-His1" 5'-Pho-GATCCgat atccaccaccaccaccaccactgaA-3' and "myco-His2" 5'-Pho-GA TCTtcagtggtggtggtggtggtgatatcG-3' (5'-phosphorylated, Metabion, Planegg-Martinsried, Germany) in vitro. For hybridization, oligonucleotides (40 μg of each) were used in a 100-μl reaction containing 10 mM Tris–HCl, pH 8.0, and 100 mM NaCl, denatured at 95 °C for 5 min, transferred to a beaker with water at 65 °C and allowed to cool to room temperature. Two microliters (1.6 μg) of this "Myco-C-His" tag were ligated to pMV262-Kan (*Bam*HI digested, dephosphorylated) to generate pSD22 or ligated to pSD24 (*Bam*HI digested, dephosphorylated) to generate pSD26. Because the hybridized "Myco-C-His" carried a 5' *Bam*HI overhang and a 3' *Bgl*II overhang (indicated by capital letters within the

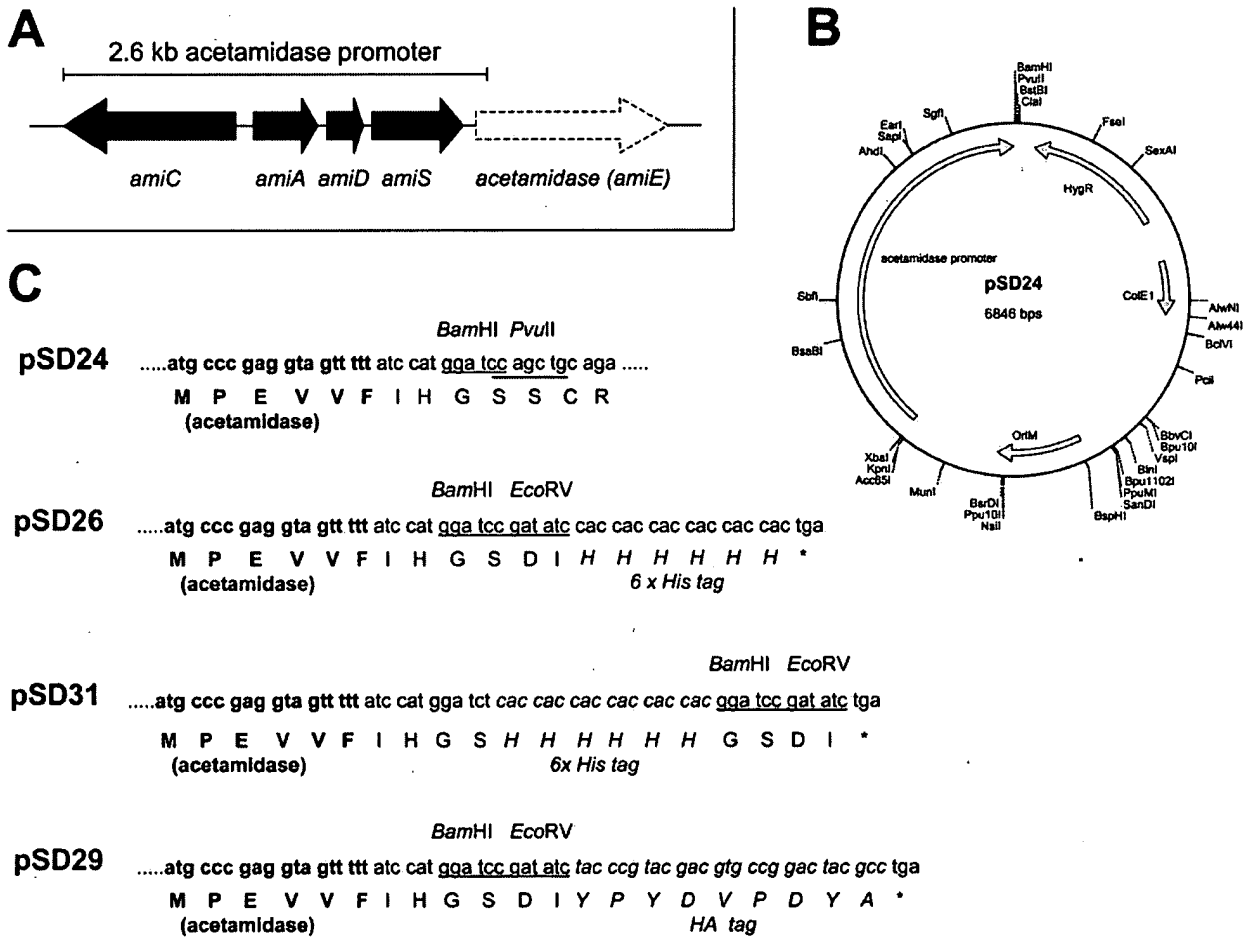

Fig. 1. *M. smegmatis* expression vectors containing the inducible acetamidase promoter. (A) The acetamidase promoter region of *M. smegmatis* (adapted from [31]). (B) Map of plasmid pSD24, an episomal *E. coli*–mycobacteria shuttle plasmid that was constructed by inserting the 2.6-kb *M. smegmatis* acetamidase promoter into the promoterless plasmid pMV206-Hygro. (C) Partial sequences of pSD24 and epitope-tagged derivatives. The unique, in-frame, *Bam*HI site of pSD24 was used to insert oligonucleotides carrying either polyhistidine- or HA-epitope tags (depicted in italics). Foreign genes can be cloned into the *Bam*HI or *Eco*RV sites of either vector (underlined), thus allowing a translational fusion with the first six amino acids of the *M. smegmatis* acetamidase (depicted in bold script).

sequence), the upstream, in-frame *Bam*HI site of either parent vector was reconstituted and the downstream *Bam*HI site was converted into a non-cleavable *Bam*HI/*Bgl*II hybrid. The "Myco-C-His" oligo also introduced a new, in-frame *Eco*RV site as well as a stop codon (Fig. 1). The double-stranded oligonucleotide "Myco-C-HA" was generated by hybridizing "myco-C-HA1" 5'-Pho-GATCCgatatctacccgtacgacgtg ccggactacgcctgaA-3' and "myco-C-HA2" 5'-Pho-GATCTtca ggcgtagtccggcacgtcgtacgggtagatatcG-3' (5'-phosphorylated, Genset, Paris, France) in vitro. Two microliters of this "Myco-C-HA" tag were ligated to pMV262-Kan (*Bam*HI digested, dephosphorylated) to generate pSD21 or ligated to pSD24 (*Bam*HI digested, dephosphorylated) to generate pSD29. The hybridized "Myco-C-HA" carried a 5' *Bam*HI overhang and a 3' *Bgl*II overhang (indicated by capital letters). Therefore, the upstream in-frame *Bam*HI of either parent vector was reconstituted and the downstream *Bam*HI site was converted into a non-cleavable *Bam*HI/*Bgl*II hybrid. The "Myco-C-HA" oligo also introduced a new, in-frame *Eco*RV site and a stop codon (Fig. 1). The double-stranded oligonucleotide "Myco-N-His" was generated by hybridizing "myco-N-His1" 5'-Pho-GATCTcaccaccaccaccaccacgga tccgatatctgaA-3' and "myco-N-His2" 5'-Pho-GATCTtcagata tcggatccgtggtggtggtggtggtgA-3' (5'-phosphorylated, MWG Biotech, Ebersberg, Germany) in vitro. The hybridized "Myco-N-His" carried *Bgl*II overhangs on each side (indicated by capital letters). Thus, both up- and downstream *Bam*HI sites were converted into non-cleavable *Bam*HI/*Bgl*II hybrids, while a new, in-frame *Bam*HI site as well as an *Eco*RV site and a stop codon were introduced (Fig. 1). Two microliters of this "Myco-N-His" tag were ligated to pSD24 (*Bam*HI digested, dephosphorylated) to generate pSD31. All epitope tags had been codon optimized for mycobacteria. Correct insertion of the oligonucleotides was verified by sequencing each plasmid with the primer "Ace-1" 5'-tcctgatcgtgtcgggcaac-3'.

2.4. *Construction of the reporter plasmids pSD24-eGFP and pMV262-eGFP*

The eGFP reporter gene was amplified from pEGFP-N1 (BD) by PCR using *Pfu*-Turbo DNA polymerase. Primers "eGFP-Start" 5'-GCGGCCGCagtgagcaagggcgaggagctg-3' and "eGFP-End" 5'-GCGGCCGCttacttgtacagctcgtccatg-3' (capital letters indicate the *Not*I restriction sites) were designed to amplify the eGFP gene without its start codon. The reading frame was corrected by including an additional adenine (indicated in bold script) at the 5'-end of eGFP, thus permitting translational fusion of eGFP with the first six amino acids of the acetamidase enzyme (Fig. 1). The 734-bp PCR product was cloned into pPCR-Script SK$^+$ Amp, using the respective cloning kit. Sequence identity was verified by sequencing the entire insert from both ends. The eGFP insert was cut out with *Not*I, followed by fill-in of 3' recessed ends with Klenow Fragment. The blunted insert was then cloned into the *Pvu*II site of either pSD24 or pMV262-Kan. Correct insertion was verified by sequencing the 5'-junction of the insert with primer "Ace-1" (see above) or "JSC77-1" 5'-gtagcggggttgccgtcacc-3' (for pMV262-Kan).

2.5. *Cloning* M. tuberculosis *ORFs*

Ten different *M. tuberculosis* ORFs and one gene-fragment were amplified by PCR using *M. tuberculosis* H37Rv chromosomal DNA as template. Specific primer pairs and DNA polymerases are listed in Table 2. The Primer3 software was used to design primers [30]. For *orf14* we used primers similar to those published previously [10]. PCR was performed in 0.2-ml thin-walled tubes employing a gradient to optimize the annealing temperature (RoboCyler Gradient96, Stratagene). PCR was carried out for one initial cycle of 5 min at 95 °C, followed by 25–30 cycles (95 °C for 1 min, gradient between 72 and 61 °C or 66 and 56 °C for 1 min, 72 °C at 3 min), and one final cycle at 72 °C for 10 min. All PCR products except for Rv3879c were blunt-end cloned into the *Sma*I site of pBluescript SK$^+$ Amp or the *Srf*I site of pPCR-Script SK$^+$ Amp. For the latter vector, the respective cloning kit was used (Stratagene). Sequence identity was verified by sequencing the first 500 bp following each junction. Rv3879c was PCR amplified with *Taq* polymerase (Invitrogen) and buffer K of the MasterAmp PCR Optimization Kit (Epicentre, Madison, WI, USA). The Rv3879c PCR product was cloned into pCR2.1-TOPO (using the TOPO TA Cloning kit, Invitrogen). In this case, cloning was done in *E. coli* Top10 (Invitrogen). Sequence identity and lack of mutations introduced by PCR were verified by sequencing the entire 2.1-kb Rv3879c insert. For insertion into expression plasmids, the *M. tuberculosis* ORFs were cut out from cloning vectors by appropriate restriction enzymes (*Bam*HI or *Bgl*II, see Table 2) and inserted into *Bam*HI-digested and dephosphorylated expression plasmids (see Table 2). Once the correct orientation of the insert had been determined by restriction analysis, *M. smegmatis* was transformed. The optimal expression vector was determined for each *M. tuberculosis* ORF: pSD26 was found to give highest expression levels for Rv3873, Rv3879c and ORF-14, while pSD31 was best for Rv3871, Rv3872, Rv3873T, CFP-10, ESAT-6, Rv3876 and Rv3878. Rv3877 could only be expressed from pSD29.

2.6. *Expression of recombinant proteins in* M. smegmatis

Starting from a single colony, a 5-ml starter culture of recombinant *M. smegmatis* carrying one of various expression plasmids was grown overnight in 7H9 complete medium with appropriate antibiotics. Starter cultures containing plasmids with the constitutive *groEL2* promoter were expanded in 7H9 complete medium and grown to stationary phase. Starter cultures containing plasmids with the inducible acetamidase promoter were used to inoculate 7H9 expression

Table 1
Strains and plasmids used in this study

Strain or plasmid	Description [a]	Source or reference
Strains		
E. coli BL21-RP (DE3)	Expression strain containing additional tRNA genes that recognize the arginine codons AGA and AGG and the proline codon CCC, respectively	Stratagene
E. coli DH5α	Standard cloning strain	Invitrogen
E. coli M15[pREP4]	Expression strain carrying the pREP4 (Kan^r) plasmid, which constitutively expresses the *lac* repressor protein, thereby regulating recombinant protein expression	Qiagen
E. coli SG13009 [pREP4]	Expression strain carrying the pREP4 (Kan^r) plasmid, which constitutively expresses the *lac* repressor protein, thereby regulating recombinant protein expression	Qiagen
E. coli Top10	Standard cloning strain	Invitrogen
M. smegmatis mc^2 155	*ept*-1	[27]
Plasmids		
pBluescript SK$^+$ Amp	Cloning vector, Amp^r	Stratagene
pCR2.1-Topo	Cloning vector, Amp^r	Invitrogen
pEGFP-N1	Cloning vector encoding the enhanced green fluorescent protein eGFP, Kan^r	BD
pET32a	Plasmid containing the inducible T7 *lac* promoter, allowing translational fusion with an N-terminal Thioredoxin (Trx) protein and an internal 6×His tag, Kan^r	Novagen
pET42a	Plasmid containing the inducible T7 *lac* promoter, allowing translational fusion with an N-terminal GST protein and an internal 6×His tag, Amp^r	Novagen
pJSC77	pMV261-Kan containing a C-terminal polyhistidine tag	[45]
pMV206-Hygro	Promoterless *E. coli*-mycobacteria shuttle vector, Hyg^r, ColE1, OriM	W.R. Jacobs Jr.
pMV262-eGFP	pMV262-Kan containing the eGFP gene, expressed as translational fusion with the first 6 amino acids of *M. bovis* BCG *groEL* 2	This study
pMV262-Kan	*E. coli*-mycobacteria shuttle vector, *groEL2* (hsp60) promoter, Kan^r, ColE1, OriM	W.R. Jacobs Jr.
pPCR-Script SK$^+$ Amp	Cloning vector for blunt ended PCR products, Amp^r	Stratagene
pQE30	Low-copy plasmid containing the inducible T5 promoter, allowing translational fusion with an N-terminal 6×His tag, Amp^r	Qiagen
pSD21	pMV262-Kan containing the "Myco-C-HA" epitope tag, allowing translational fusion with a C-terminal HA tag	This study
pSD22	pMV262-Kan containing the "Myco-C-His" epitope tag, allowing translational fusion with a C-terminal 6×His tag	This study
pSD24	*E. coli*-mycobacteria shuttle vector, Hyg^r, ColE1, OriM, contains the acetamidase promoter	This study
pSD24-eGFP	pSD24 containing the eGFP gene, expressed as translational fusion with the first six amino acids of the acetamidase enzyme	This study
pSD26	pSD24 containing the "Myco-C-His" epitope tag, allowing translational fusion with a C-terminal polyhistidine tag	This study
pSD29	pSD24 containing the "Myco-C-HA" epitope tag, allowing translational fusion with a C-terminal HA tag	This study
pSD31	pSD24 containing the "Myco-N-His" epitope tag, allowing translational fusion with an N-terminal polyhistidine tag	This study

[a] Amp^r, ampicillin resistance; Kan^r, kanamycin resistance; Hyg^r, hygromycin resistance.

medium with appropriate antibiotics at a ratio of 1:50, and these expression cultures were grown overnight. When they reached OD_{600} = 0.6, a small sample was transferred to another flask and cultured separately as "non-induced control". The remaining culture was induced with 0.2% acetamide (Sigma A-0500; used as a 44% stock solution in H_2O) and grown for another 4–7 h. Bacteria were pelleted, washed once in PBS–0.05% Tween-80 and resuspended in Buffer B (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris/HCl, pH 8.0) at 1/100 of the original volume. Then, 1% (v/v) of a protease inhibitor cocktail (Sigma P-8340) was added. Bacteria were disrupted by sonication in a chilled cup-sonicator (Branson Sonifier 450) for 10 min at maximum output and 50% duty cycle. The crude extract was transferred to microcentrifuge tubes and clarified by spinning for 10 min at 13 000 rpm in a precooled centrifuge.

2.7. Promoter activity tests

A comparative experiment was set up in which *M. smegmatis* carrying pMV262-eGFP or pSD24-eGFP was grown in 100 ml of 7H9 complete medium with appropriate antibiotics at 37 °C until OD_{600} = 0.6 was reached. The pSD24 vector contains the inducible acetamidase promoter. The pSD24-eGFP cultures were set up in duplicate with one culture being induced with 0.2% acetamide and one left untreated. The plasmid pMV262-eGFP contains the *groEL2* promoter, therefore, the culture was left untreated. At hourly intervals, 11-ml samples were removed, of which 1 ml was used to measure the OD_{600} and the remaining 10 ml were kept at 4 °C until all samples were collected. At the end of the experiment, 1 ml of 10% paraformaldehyde was added to each 10-ml sample in order to fix bacteria. Fixed bacterial

Table 2
Cloning of *M. tuberculosis* ORFs

ORF	Primer [a]	DNA polymerase	Cloning vector	Number of cloned nucleotides (% of Orf)	Vectors used for expression
Rv3871	3871–START: 5'-GGATCCatgactgctgaaccggaag-3' 3871–END: 5'-GGATCCaccggcgcttgggggtgc-3'	Pfu Turbo	pBluescript	1773 (100)	pQE30, pSD22, pSD26, pSD31
Rv3872	3872–START2: 5'-GGATCCgacattggcacgcaagtgag-3' 3872–END2: 5'-GGATCCgtcgtcgatttgcgaataggt-3'	Pfu Turbo	pPCR-Script	240 (81)	pQE30, pSD26, pSD31
Rv3873	3873–START: 5'-GGATCCatgctgtggcacgcaatg-3' 3873–END: 5'-GGATCCccagtcgtcctcttcgtc-3'	Pfu Turbo	pBluescript	1104 (100)	pQE30, pSD21, pSD22, pSD26
Rv3873T	3873T–START: 5'-GGATCCacaccggttggccagttg-3' 3873–END: 5'-GGATCCccagtcgtcctcttcgtc-3'	Pfu Turbo	pPCR-Script	519 (47)	pSD26, pSD31
Rv3874, *lhp*	3874–START: 5'-GGATCCatggcagagatgaagacc-3' 3874–END: 5'-GGATCCgaagcccatttgcgaggac-3'	Pfu-Turbo	pBluescript	300 (100)	pQE30, pET32a, pET42a, pSD21, pSD22, pSD26, pSD31
Rv3875, *esat-6*	esat6–START: 5'-GGATCCatgacagagcagcagtggaatttc-3' esat6–END: 5'-GGATCCtgcgaacatcccagtgacgttg-3'	Pfu-Turbo	pPCR-Script	285 (100)	pQE30, pJSC77, pSD22, pSD26, pSD31
Rv3876	3876– *Bgl* II-START: AGATCTatggcggccgactacgacaagctcttccgg-3' 3876– *Bgl* II-END: AGATCTacgacgtccagccctctc-3'	Pfu-Turbo	pPCR-Script	1998 (100)	pJSC77, pSD26, pSD31
Rv3877	3877–START: 5'-GGATCCttgagcgcacctgctgttg-3' 3877–END: 5'-GGATCCgaaccggatattgcggac-3'	Pfu-Turbo	pBluescript	1533 (100)	pET42a, pSD21, pSD22, pSD26, pSD29, pSD31
Rv3878	3878–START: 5'-GGATCCatggctgaaccgttggcc-3' 3878–END: 5'-GGATCC-caacgttgtggttgttgag-3'	Pfu-Turbo	pBluescript	840 (100)	pQE30, pSD22, pSD26, pSD31
Rv3879c	3879c–START2: 5'-AGATCTggcagctatgccagacagatg-3' 3879c–END2: 5'-AGATCTgagcaacccggtgacgtattg-3'	Taq	pCR2.1-Topo	2142 (98)	pSD26, pSD31
orf14	ORF14–START: 5'-GGATCCatgctcggtgccgccgccttg-3' ORF14–END: 5'-GGATCCgcagaagtcgccgccaccccc-3'	Pfu-Turbo	pPCR-Script	735 (100)	pSD22, pSD26

[a] Capital letters indicate the added restriction sites.

cultures were diluted 1:100 in sterile-filtered PBS-0.05% Tween-20, and expression of eGFP was determined in a flow cytometer (FACScalibur, BD). Gating parameters were determined by discriminating against background fluorescence and collecting only FL-1-positive cells. Analysis of eGFP expression was performed with the WinMDI 2.8 software (http://facs.scripps.edu). Mycobacteria tend to clump; therefore, a gate was set that comprised single bacteria but excluded larger aggregates. The median of fluorescence within this gate was determined.

2.8. Immunoblotting

Proteins were separated by conventional one-dimensional SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Standard minigels of 1–1.5-mm thickness and 10–15% acrylamide density were prepared using the Mini-Protean III system (Bio-Rad, Hercules, CA, USA). Rainbow markers of two different size ranges were used (Amersham Biosciences, Freiburg, Germany). Protein gels were either stained with Coomassie Brilliant Blue (Sigma) or transferred to nitrocel-

Table 3
Purification of recombinant *M. tuberculosis* RD1 proteins

Protein	Purification	Identity confirmed [a]	Molecular mass	
			Estimated [b]	Observed [c]
Rv3871	Ni^{2+} chelator → Gel filtration → MonoQ	Yes (50% s.c.) [d]	64.6	70
Rv3872	Co^{2+} chelator → Gel filtration	Yes	9.9	10
Rv3873	Co^{2+} chelator → RP-HPLC (C18)	Yes (32% s.c.)	37.3	45
Rv3873T	Co^{2+} chelator → Gel filtration	Yes	19.9 [d]	22
CFP-10	Co^{2+} chelator → Anion exchange	Yes (85% s.c.)	10.8	14
ESAT-6	Co^{2+} chelator → Anion exchange → Cation exchange	Yes (45% s.c.)	9.9	10
Rv3876	Ni^{2+} chelator → MonoQ → Gel filtration	Yes (55% s.c.)	70.7	110
Rv3877	n.d.	n.d.	54.0	60 + 100
Rv3878	Co^{2+} chelator → Anion exchange	n.d.	27.4	40
Rv3879c	Co^{2+} chelator → MonoQ → Gel filtration → MiniQ	Yes (45% s.c.)	74.5	100
ORF-14	Co^{2+} chelator → Gel filtration	Yes	28.2 [e]	35

[a] Identity confirmed by MALDI-MS PMM.
[b] TubercuList.
[c] Observed in 1D-PAGE.
[d] Sequence coverage.
[e] Number of amino acids × 115.

lulose membranes (Amersham) by semi-dry blotting techniques. Transferred proteins were stained for 5 min with Ponceau S (Serva) and destained with washing buffer (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.1% Tween-20 (Sigma)). Membranes were blocked with 1% BSA in washing buffer at 4 °C overnight. The primary antibody was added for 1 h at RT, and afterwards membranes were washed several times with washing buffer. Freeze-dried mouse anti-Penta-His monoclonal antibody (Qiagen, Hilden, Germany) was reconstituted to 200 μg/ml and used at 1:2000 to 1:5000 dilution in washing buffer. Mouse-anti-GFP monoclonal antibody (BD) was diluted 1:10 000 in washing buffer. Monoclonal mouse anti-HA antibody (BAbCo, Richmond, CA, USA) was diluted 1:10 000 in washing buffer. Monoclonal anti-ESAT-6 antibody HYB-76 (kind gift of Peter Andersen, Statens Serum Institute, Copenhagen, Denmark) was used at 1:200 dilution. Sera from 11 sputum-positive tuberculosis patients (kind gift of Karlheinz Neumann, Zentralklinik Emil von Behring, Berlin, Germany. Patient sera obtained by informed consent.) were pooled and diluted 1:200 in washing buffer. Sera from female BALB/c mice were taken 11 weeks post i.v. infection with 6.8×10^6 *M. tuberculosis* H37Rv (kindly provided by Peter Aichele and Peter Seiler, Max Planck Institute for Infection Biology, Berlin, Germany) and diluted 1:1000 in washing buffer. Peroxidase-conjugated goat anti-mouse-IgG (H + L) antibody (Jackson ImmunoResearch, West Grove, PA, USA) was diluted 1:20 000 in washing buffer. Peroxidase-conjugated goat anti-human-IgG (H + L) antibody (Jackson) was diluted 1:50 000 in washing buffer. The secondary antibodies were added for 1 h at RT, and blots were developed with the ECL system (Amersham) after extensive washing.

2.9. Protein purification

Protein extracts were clarified by centrifugation in a Sorvall SS34 rotor (18 000 rpm, 1 h). Capture of proteins was done on a metal chelating sepharose column (size: 26 mm i.d., 50 mm length) charged with $NiCl_2$ or $CoCl_2$ (see Table 3). Chromatography was performed according to standard routines for metal chelating purification under denaturating conditions in the presence of 8 M urea, and elution was achieved with a linear gradient of 0–500 mM imidazole in buffer B over 100 ml at a flow rate of 1 ml/min. Purity of eluted proteins was assessed by SDS-PAGE and immunoblotting analysis. Polishing of proteins required further chromatographic steps as listed in Table 3. Gelfiltration was done on a Pharmacia Superdex HR200 (HR10/30) column or on a Superose 12 (HR10/30) column in buffer B with 200 mM NaCl at a flow rate of 0.4 ml/min. Before loading protein samples onto the column, they were concentrated to 0.5 ml by ultrafiltration (Amicon Ultrafree 15). For ion exchange, proteins were dialyzed against 8 M urea buffered with 20 mM Tris–HCl at pH 8.0 or with 40 mM MES at pH 6.0, depending on the chromatography used. Anion exchange chromatography was performed on either a MonoQ (HR5/5) column, a MiniQ column (Pharmacia) or a POROS HQ20 column (4.6 × 100 mm, Perseptive). Proteins were eluted with a linear gradient of NaCl. Cation exchange was done on a POROS HS20 column (4.6 × 100 mm). For reversed phase HPLC, proteins were concentrated by ultrafiltration to 1 ml and directly loaded on a Prontosil AQ C18 column (4.0 × 100 mm, Bischoff chromatography), which was equilibrated with 0.1% TFA and 5% acetonitrile. Proteins were eluted with a linear gradient of increasing acetonitrile concentration in the presence of 0.1% TFA. Proteins were lyophilized and dissolved in buffer B.

3. Results

3.1. Construction of epitope-tagged expression plasmids

Based on previous studies [28,29,31], primers were designed to amplify the acetamidase promoter from *M. smegmatis* mc^2155 chromosomal DNA, including all potentially

regulatory ORFs (*amiC, amiA* and *amiD*). The 2.6-kb acetamidase promoter was inserted into pMV206, a promoterless episomal *E. coli*–mycobacteria shuttle plasmid carrying a hygromycin-resistance gene (Table 1). The resulting 6.8-kb vector, pSD24, allows expression of foreign genes by translational fusion with the first six amino acids of the *M. smegmatis* acetamidase AmiE (Fig. 1). The plasmid pSD24 was used to construct a family of epitope-tagged versions: a C-terminal polyhistidine (6×His) tag was inserted to create pSD26, an N-terminal 6×His-tag was inserted to construct pSD31, and a C-terminal influenza hemagglutinin (HA)-tag was introduced to create pSD29 (Fig. 1). The sequence of both epitope tags had been codon optimized for expression in mycobacteria. In the same manner, epitope tags were inserted into pMV262, an episomal *E. coli*–mycobacteria shuttle plasmid containing the constitutive *groEL2* promoter and a kanamycin-resistance gene, thus constructing pSD21 (C-terminal HA-tag) and pSD22 (C-terminal 6×His-tag). All epitope-tagged vectors were designed to carry two convenient restriction sites, *Bam*HI and *Eco*RV. Insertion of foreign genes into either site allowed in-frame translational fusion with the first six amino acids of *M. smegmatis* acetamidase (for pSD26, pSD29 and pSD31) or the first six amino acids of *M. tuberculosis groEL2* (for pSD21 and pSD22) as well as with respective tags.

3.2. Expression of a green fluorescent reporter protein

The eGFP gene was inserted into pSD24 or pMV262 to create reporter plasmids to compare the promoter strength of the constitutive *groEL2* promoter with the promoter strength of the inducible acetamidase promoter. The reporter plasmids were introduced into *M. smegmatis*, and cultures were grown until $OD_{600} = 0.6$ was reached. This time point was defined as $t = 0$. Depending on the promoter, cultures were then induced with 0.2% acetamide or left untreated (see Section 2). Expression of eGFP in *M. smegmatis* was determined by analysis of identical cell numbers in a flow cytometer (FACS). Fig. 2B shows that bacterial growth was not influenced by addition of acetamide. In *M. smegmatis*, the *groEL2* promoter was already "on" at $t = 0$, and constitutive expression of eGFP remained almost constant during the 7 h of the experiment (Fig. 2A). In contrast, 2 h were required after induction of the acetamidase promoter until expression of eGFP became detectable. Thereafter, expression continuously increased and reached a maximum at 7 h after induction (latest time point of the experiment). The acetamidase promoter appears to be tightly regulated. Without induction, no expression of eGFP was visible by FACS analysis (Fig. 2A). Neither of the two mycobacterial promoters allowed significant expression of eGFP in *E. coli* (data not shown). Concluding from these experiments, the acetamidase promoter appears to be two- to eight-fold stronger than the *groEL2* promoter, depending on duration of protein expression.

Previous studies on the inducible nature of the acetamidase promoter employed rather sophisticated culture media

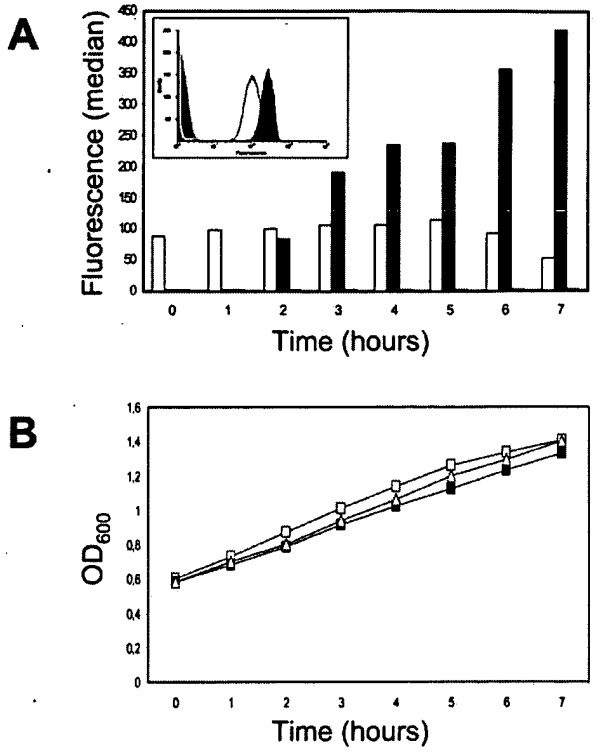

Fig. 2. Bacterial growth and promoter-dependent expression of eGFP. *M. smegmatis* mc^2155 was transformed with pMV262–eGFP (containing the constitutive *groEL2* promoter) or pSD24-eGFP (containing the inducible acetamidase promoter). Cultures were grown to an $OD_{600} = 0.6$ (t_0) and left untreated or induced with 0.2% acetamide. (A) Promoter-dependent expression of eGFP as measured by FACS analysis (open white bars, pMV262-eGFP; black bars, pSD24-eGFP, induced; gray bars, pSD24-eGFP, not induced). The insert presents an overlay of three individual histograms at $t = 4$ h. (B) Bacterial growth (□, pMV262-eGFP; ■, pSD24-eGFP, induced; Δ, pSD24-eGFP, not induced).

and induction procedures [12,28,29,32]. We wanted to simplify the acetamidase promoter system by using standard 7H9 mycobacterial culture medium containing glucose as carbon source. *M. smegmatis* carrying the reporter plasmid pSD24-eGFP was grown in either Middlebrook 7H9 "complete medium" (supplemented with 10% ADS) or protein-free Middlebrook 7H9 "expression medium" (see Section 2). Fig. 3A shows that similar amounts of eGFP were produced after induction with 0.2% acetamide irrespective of the supplementation of the culture medium with ADS. The immunoblot shown in Fig. 3B confirmed the identity of the prominent 27-kDa protein as eGFP. This experiment strengthened our previous observation about the tight regulation of the acetamidase promoter (Fig. 2) by demonstrating that in the absence of the inducing agent only a small amount of eGFP was detectable.

3.3. Cloning of the entire gene set encoded within the RD1 region of M. tuberculosis

The 9455-bp RD1 of *M. tuberculosis* is absent from all attenuated *M. bovis* BCG vaccine strains due to a natural deletion event [6,7]. Despite their potential role in pathogen-

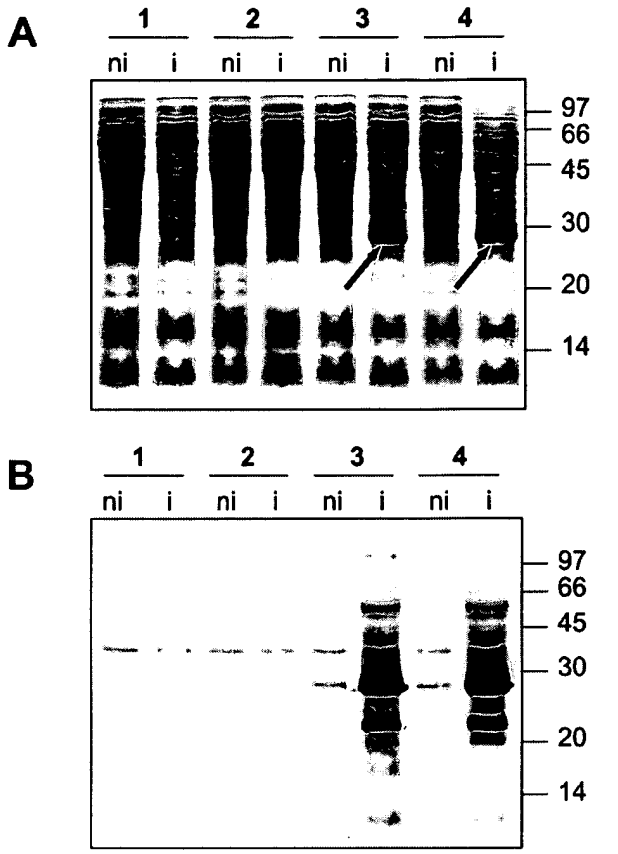

Fig. 3. Expression of eGFP in protein-free medium. *M. smegmatis* mc^2155 was transformed with the control plasmid pSD24 (lanes 1 + 2) or pSD24-eGFP (lanes 3 + 4). Cultures were grown either in 7H9 complete medium (lanes 1 + 3) or protein-free 7H9 expression medium (lanes 2 + 4) to an $OD_{600} = 0.6$. Then, half of each culture was induced with 0.2% acetamide (i), while the other half was left untreated, i.e. not induced (ni). Bacteria were harvested 4 h after induction. Bacterial extracts were subjected to 15% SDS-PAGE, and proteins were stained with Coomassie brilliant blue (A) or transferred to nitrocellulose membranes and stained with anti-GFP monoclonal antibody (B). The arrows mark the position of the eGFP protein; molecular weight standards (in kDa) are indicated on the right.

esis and virulence, only two of the RD1 proteins, namely CFP-10 and ESAT-6, have been studied in detail. Our aim was to clone, express and study the entire set of gene products of the *M. tuberculosis* RD1 region (Fig. 4), primarily because we are interested in its diagnostic potential. Primer pairs were designed to amplify the entire ORF or—if unfeasible—as much coding sequence as possible of Rv3871–Rv3879c, the nine genes that are found within the RD1 region, according to the annotation by the genome project (see Table 2). For Rv3873, a second set of primers was designed to amplify only the last 519 bp of this gene. This ORF was named Rv3873T, because it codes for a truncated version of this PPE protein. Rv3873T represents the specific C terminus but not the highly conserved N-terminal domain. Besides the nine annotated RD1 genes, we also cloned *orf14*, as it had been shown to encode for an immunoreactive protein [10]. All PCR products were generated with proofreading DNA polymerases and cloned into pBluescript or similar cloning vectors (Table 2). The PCR product for Rv3979c could only be generated by using *Taq* polymerase and an optimized buffer. After sequence verification, individual RD1 ORFs were subcloned into various expression vectors.

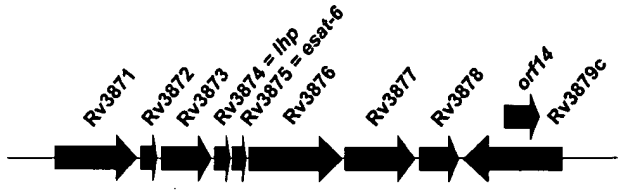

Fig. 4. The RD1 region of *M. tuberculosis*. Ten of the ORFs encoded within the RD1 region are shown as black arrows. The length, orientation and spacing of the arrows indicates the approximate location of the ORFs within the RD1 region.

3.4. Expression and purification of RD1 proteins

Initially, we attempted to express seven of the *M. tuberculosis* RD1 proteins in *E. coli* by using standard vector systems (Table 2). After testing several expression vectors and various *E. coli* host strains (including one that was codon optimized for GC-rich organisms), we could express four recombinant RD1 proteins (Rv3871, CFP-10, ESAT-6 and Rv3878) without problems, while the remaining three proteins (Rv3872, Rv3873 and Rv3877) could not be expressed at all or only at very low levels (data not shown). Since we were determined to produce the entire set of RD1 proteins for further studies, we decided to discontinue the *E. coli* system and to express *M. tuberculosis* proteins in *M. smegmatis*. Two episomal *E. coli*–mycobacteria shuttle plasmids, pSD21 and pSD22, were generated. Both vectors contained the constitutive *groEL2* promoter and a C-terminal HA- or polyhistidine tag for identification and purification of the recombinant protein (described above in Section 3.1). Several of the *M. tuberculosis* RD1 ORFs were inserted into pSD21 or pSD22 (see Table 2), and in some cases expression was achieved (data not shown), but the system proved unsatisfactory. The two main reasons were toxicity of recombinant proteins due to constitutive expression or overall low protein yields. Similar problems have been described for expression of a PPE protein under control of the *groEL2* promoter [12]. These failures prompted the development of the vectors carrying the inducible acetamidase promoter described in Section 3.1. All RD1 ORFs were subcloned into one or several of the vectors (Table 2), and the resulting plasmids were introduced into *M. smegmatis*. We found that the position of the 6×His tag was important for expression, immunodetection and binding of recombinant proteins to metal chelating matrices. Fig. 5 shows that by using an optimal vector, all 11 RD1 proteins could be expressed within 4 h after induction with 0.2% acetamide, albeit at different levels. Non-induced controls were always negative for the respective proteins (data not shown). Expression of most recombinant *M. tuberculosis* RD1 proteins was clearly visible by Coomassie staining, with the exception of Rv3876 and Rv3877.

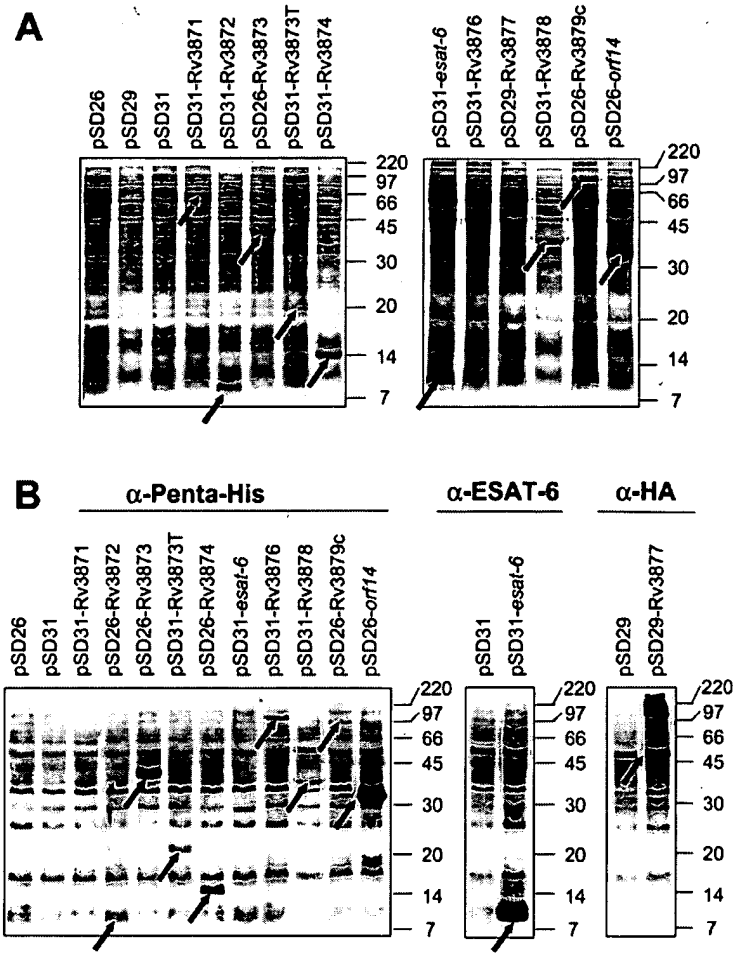

Fig. 5. Expression of 11 RD1 proteins in *M. smegmatis*. Cultures of *M. smegmatis* mc^{2}155 transformed with indicated plasmids were grown in 7H9 expression medium to an OD_{600} = 0.6. Then, each culture was induced with 0.2% acetamide, and bacteria were harvested 4 h after induction. Bacterial extracts were subjected to 15% SDS-PAGE, and proteins were stained with Coomassie brilliant blue (A) or transferred to nitrocellulose membranes and stained with anti-Penta-His, anti-HA or anti-ESAT-6 monoclonal antibodies, as indicated (B). The arrows mark the position of the induced RD1 proteins; the position of molecular weight standards (in kDa) is shown on the right. Please note that expression vectors vary for Rv3872 and Rv3874 in (A) and (B).

However, these proteins became detectable after immunoblotting with anti-His or anti-HA monoclonal antibodies. In contrast, Rv3871 and ESAT-6 could be seen in Coomassie gels but could not be stained with anti-His antibodies. On the other hand, a monoclonal antibody to ESAT-6 reacted strongly with the recombinant protein. In general, the pSD31 vector which allowed translational fusion with an N-terminal 6×His tag was superior to the pSD26 (C-terminal 6×His tagged) vector. Toxicity problems were not encountered for any of the RD1 proteins, except for the probable membrane protein Rv3877, which is predicted to express 11 transmembrane domains [26]. We could not express Rv3877 with a polyhistidine tag at either end, but expression was achieved with a C-terminal HA tag (Fig. 5B). No expression problems were observed for very small proteins (e.g. the 10-kDa protein Rv3872) or very large proteins (e.g. Rv3876 and Rv3879c with apparent molecular masses >100 kDa). Recombinant proteins were stable even when expression was carried out for up to 22 h (data not shown). Initial expression experiments were performed in 50-ml cultures. Once the optimal vector had been determined, cultures were scaled up to 1–3 l, and recombinant proteins were purified by conventional chromatographic methods (Table 3). Fig. 6A shows that only four of the RD1 protein bands were found to migrate at their estimated molecular mass or up to 10% larger (Rv3871, Rv3872, Rv3873T and ESAT-6), three proteins ran up to 30% larger (Rv3873, CFP-10 and ORF-14), while Rv3876, Rv3878 and Rv3979c appeared 30–60% larger than estimated (see Table 3). This phenomenon may partially be attributed to posttranslational modifications or—more likely—to the unusual amino acid composition of some proteins (e.g. Rv3876, Rv3878 and Rv3879c, which are rich in proline and alanine), a fact that has been observed before for the proline-rich Apa protein [33]. In our hands the recombinant ORF-14 protein was clearly visible in Coomassie gels, while Ahmad et al. [10] found that pure ORF-14 protein (produced by *E. coli*) did not bind Coomassie. The truncated Rv3873T protein did not stain well with Coomassie but was detected by anti-His monoclonal antibody, as were all other RD1 proteins except for Rv3871 and ESAT-6 (Fig. 6B).

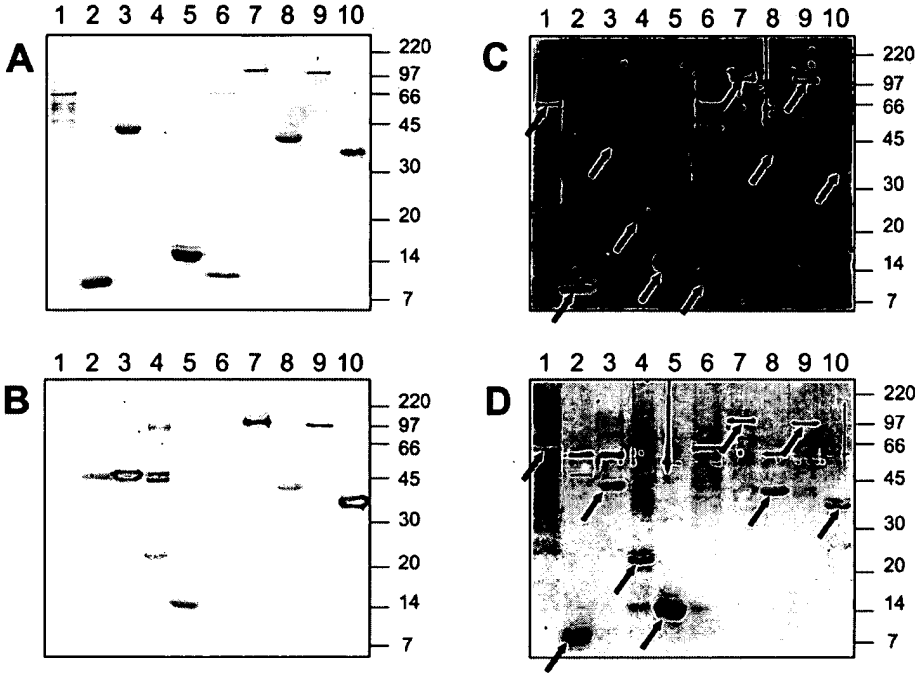

Fig. 6. Detection of antibodies to RD1 proteins in human and murine sera. Purified RD1 proteins were loaded on 15% SDS-PAGE gels (3–20 μl each; the volume was adjusted to give bands of uniform intensity). Lane numbers indicate: (1) Rv3871, (2) Rv3872, (3) Rv3873, (4) Rv3873T, (5) CFP-10, (6) ESAT-6, (7) Rv3876, (8) Rv3878, (9) Rv3879c, (10) ORF-14. The position of the molecular weight standards (in kDa) is indicated on the right. (A) Coomassie brilliant blue stained gel showing purified RD1 proteins. (B) RD1 proteins detected by anti-Penta-His monoclonal antibody. (C) Recognition of RD1 proteins by pooled sera from sputum-positive tuberculosis patients. The arrows mark the position of the RD1 proteins (D) Recognition of RD1 proteins by antibodies from *M. tuberculosis* H37Rv-infected mice.

Identity of all RD1 proteins was confirmed by subjecting purified proteins to mass spectrometry (after separation on SDS-PAGE gels, staining with Coomassie brilliant blue and tryptic digestion of respective protein bands; see Table 3).

3.5. Expression of RD1 proteins in vivo

Except for CFP-10 and ESAT-6, which are strong T- and B-cell antigens, it is presently not known if any of the other *M. tuberculosis* RD1 proteins are expressed in vivo. Therefore, immunoblots were performed to detect antibodies to *M. tuberculosis* RD1 proteins in human and murine sera. Fig. 6C shows that pooled sera from 11 sputum-positive tuberculosis patients reacted with all RD1 proteins, although responses to Rv3873, Rv3873T and ESAT-6 were weak. Fig. 6D demonstrates that pooled sera from *M. tuberculosis*-infected BALB/c mice contained antibodies to all RD1 antigens except for ESAT-6. However, it is possible that ESAT-6 is recognized by murine sera at earlier or later time points after infection. It is apparent that human and murine sera also recognized impurities in protein preparations that were invisible in Coomassie gels. The fact that antibodies against RD1 proteins were detectable in human and murine sera suggests the conclusion that all RD1 proteins are expressed by *M. tuberculosis* in vivo and underscores the possibility that the RD1 gene products play a role in virulence and pathogenesis.

3.6. Glycosylation of RD1 proteins

Several *M. tuberculosis* and *M. bovis* proteins including the 19-kDa lipoprotein are known to be glycosylated [34–36]. *M. smegmatis* is able to glycosylate recombinant

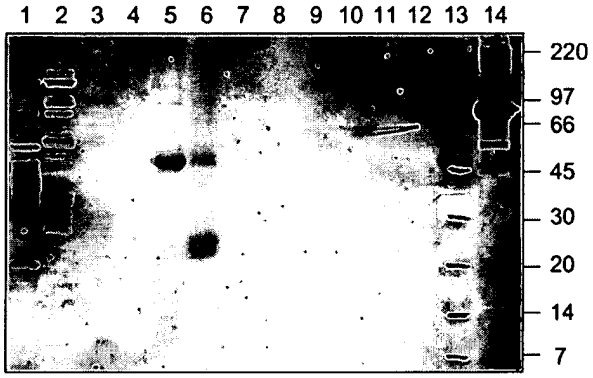

Fig. 7. Glycosylation of RD1 proteins produced in *M. smegmatis*. Protein extracts from *M. tuberculosis* H37Rv (lane 1), *M. smegmatis* (lane 2), purified RD1 proteins ((3) Rv3871, (4) Rv3872, (5) Rv3873, (6) Rv3873T, (7) CFP-10, (8) ESAT-6, (9) Rv3876, (10) Rv3878, (11) Rv3879c, (12) ORF-14) and transferrin as positive control (lane 14) were loaded on 15% SDS-PAGE gels (3–20 μl each; the volume was adjusted to give bands of uniform intensity) and transferred to nitrocellulose membranes. Low-molecular-weight markers were loaded in lane 13; the position of high-molecular-weight markers (in kDa) is indicated on the right. Glycoproteins were identified by using the DIG Glycan Detection kit (Roche), according to the manufacturer's instructions.

M. tuberculosis 19-kDa protein [37]. Except for ESAT-6 (which is not glycosylated [19]), it is currently not known whether the *M. tuberculosis* RD1 proteins are naturally glycosylated. We, therefore, analyzed the glycosylation status of the *M. tuberculosis* RD1 proteins produced in recombinant form in *M. smegmatis* and found that the PPE protein Rv3873 as well as the truncated form Rv3873T stained positive (Fig. 7). Although this is a preliminary result, it indicates that the glycosylation site(s) are not on the cross-reactive, highly conserved N-terminal domain but that the specific C-terminal portion of this PPE protein carries the glycosylation site(s). None of the remaining RD1 proteins appeared to be glycosylated. Fig. 7 also shows that several bands in *M. tuberculosis* and *M. smegmatis* protein extracts stained positive for glycosylation but that the patterns were distinctly different between the two strains.

4. Discussion

Since the completion of the *M. tuberculosis* genome in 1998 [1], high-tech applications such as genomics, transcriptomics and proteomics have become feasible which will promote deeper insights into host–pathogen interactions. However, the mundane mission to study the function of *M. tuberculosis* proteins is still hampered by the difficulties in expressing such proteins in recombinant form. The high GC content of *M. tuberculosis* DNA, uncommon codon usage and preferences for certain amino acids may cause problems for overexpression in conventional *E. coli* systems. For the time being, successful expression is highly arbitrary, but in the future, bioinformatics may be able to predict challenging proteins and offer solutions. A more biological approach is to engineer non-pathogenic mycobacteria to express foreign proteins, thus making use of an autologous host system. Based on the construction of *E. coli*–mycobacteria shuttle plasmids [38,39], a large number of recombinant *M. bovis* BCG strains have been created and used for vaccination purposes [40], while saprophytic, fast-growing *M. smegmatis* organisms have been employed for expression of *M. tuberculosis* proteins, yet mainly for studying mycobacterial genetics. Although a variety of constitutive promoters and signals have been used to express *M. tuberculosis* proteins, toxicity problems and low protein yields prevailed. It was not until the inducible acetamidase promoter had been cloned and characterized [28,29,31] that *M. smegmatis* could be successfully used for high-yield production of *M. tuberculosis* and *Mycobacterium leprae* proteins [12,32].

In this study we created a family of expression vectors carrying the inducible acetamidase promoter. The acetamidase promoter was found to be inactive in *E. coli* ; consequently, foreign genes can be inserted into vectors carrying this promoter without expression of the potentially toxic gene product. We confirmed that the acetamidase promoter is tightly regulated in *M. smegmatis* [32] and showed that it is about two- to eight-fold stronger than the constitutive *groEL2* promoter. Similar expression vectors have been described before, but they were not suited for our purposes, as they lacked suitable restriction sites or epitope tags, or some of the regulatory ORFs of the acetamidase promoter were missing [28,32]. Our vectors carry two convenient restriction sites, *Bam*HI and *Eco*RV, which allow insertion of foreign genes. Most *M. tuberculosis* ORFs can be cloned into the *Bam*HI site, as they rarely contain endogenous *Bam*HI recognition sequences, but should this be the case, *Bgl*II linkers can be used. Alternatively, any blunt-end cutting restriction enzyme can be employed to insert ORFs into the *Eco*RV site. We have not used directional *Bam*HI–*Eco*RV cloning, as the two sites are very close to each other, but this should be possible. Our expression system proved to work well in standard mycobacterial 7H9 culture medium. Thus, there was no need to employ the rather sophisticated culture media and induction procedures used previously [12,28,29,32]. The system is also very robust, as the entire set of the 11 *M. tuberculosis* RD1 proteins (comprising proteins from 95 to 729 aa in length) could be expressed. Although an optimal vector had to be determined for each candidate, all recombinant *M. tuberculosis* RD1 proteins were produced within only 4 h (i.e. 1.3 generation times for *M. smegmatis*). Our vectors might also function in slow-growing mycobacteria, since similar plasmids bearing the inducible acetamidase promoter have been used to express recombinant proteins in *M. tuberculosis* and *M. bovis* BCG [12,41]. However, we have not been able to transform *M. bovis* BCG with pSD24, while the parent plasmid pMV206 transformed well (data not shown). It appears that the acetamidase promoter cassette had been inserted into a region of the plasmid which interfered with replication, but this problem may be remedied by subcloning the cassette into another site.

After purifying the recombinant *M. tuberculosis* RD1 proteins, identity was confirmed by MALDI-MS. Only the putative membrane protein Rv3877, which is predicted to contain 11 transmembrane domains [26], could not be purified. Expression of this protein, however, became possible with a C-terminal HA-tag (although most protein appeared in high-molecular-weight aggregates, see Fig. 5B), while expression proved to be toxic for *M. smegmatis* when a polyhistidine tag was added to either end (data not shown). Denaturing conditions were required to solubilize the aggregates, but this treatment prevented binding to an affinity column bearing anti-HA monoclonal antibody (data not shown). The inducible acetamidase promoter system will be useful for the expression of membrane proteins, but more work is required to ensure solubility and correct folding of recombinant proteins and to establish suitable purification procedures. When CFP-10 and ESAT-6 were purified, we found that CFP-10 was negatively charged and strongly bound to an anion exchange column, while ESAT-6 did not bind to the same matrix but was identified in the flow-through. We consider this interesting because it has recently been described that the genes encoding CFP-10 and ESAT-6 are organized in an operon and that the gene products form tight 1:1 complexes [17,23]. We also found that purified

ESAT-6 formed homo-dimers and -trimers, which is surprising, because the protein was stored in 8 M urea-phosphate buffer, and SDS-PAGE gels were run under denaturing conditions (data not shown). It is, therefore, tempting to speculate that opposite charges contribute to formation and/or stability of the CFP-10/ESAT-6 complexes and that—without a natural partner—at least ESAT-6 can form multimers. By using sera from tuberculosis patients and *M. tuberculosis*-infected mice, we showed that all RD1 proteins (except for Rv3877, which was not included in the test) were expressed in vivo. Some of the RD1 proteins have been used in diagnostic tests before, and it is known that *E. coli*-derived Rv3871, Rv3872, Rv3873, CFP-10, ESAT-6 and Rv3878 elicit delayed-type hypersensitivity reactions in *M. tuberculosis*-infected guinea pigs and that the antigens are recognized by human sera, i.e. expressed in vivo or bearing cross-reactive domains [15–17,21,42]. In vivo expression of ORF-14 was also noted before [10], but expression of Rv3879c, a protein of unknown function and with no other mycobacterial homologues, has not been described before.

Glycosylation is a posttranslational modification, which is important for immune responses to mycobacterial antigens [33]. Apart from the aforementioned obstacles to expressing mycobacterial proteins in *E. coli, M. tuberculosis* antigens derived from a recombinant *E. coli* host have been shown to be inferior to antigens produced by recombinant *M. smegmatis* [43,44]. In this study we present evidence that the *M. tuberculosis* PPE protein Rv3873 is glycosylated at its C-terminal domain. Glycosylation does not seem to be a general feature of the PPE proteins, since another member, Rv1917c, was not found to be glycosylated, although it had also been produced in recombinant *M. smegmatis* [12].

In conclusion, a robust and inducible expression system for *M. tuberculosis* proteins in the fast-growing strain *M. smegmatis* was described. Ten *M. tuberculosis* proteins encoded by the RD1 region were produced, purified and shown to be expressed in vivo. The system may even be able to express membrane proteins. The PPE protein Rv3873 was shown to be glycosylated at its C-terminal domain. We hope that our protein expression system will facilitate the study of many more *M. tuberculosis* proteins.

Acknowledgements

Many thanks to Peter Aichele, Peter Andersen, William R. Jacobs Jr., Peter Seiler and Karlheinz Neumann for providing reagents, to Olaf Mertsch for technical assistance and to Volker Pfeifer and Martin S. Pavelka Jr. for critical reading of the manuscript and advice. This work was supported by a postdoctoral fellowship from the German government "Infektionsforschung und AIDS-Stipendienprogramm" to S. Daugelat, and by BMBF "Optimized Proteome Analysis" to S. Daugelat and S.H.E. Kaufmann.

References

[1] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.

[2] J.C. Camus, M.J. Pryor, C. Medigue, S.T. Cole, Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv, Microbiol. 148 (2002) 2967–2973.

[3] R. Brosch, S.V. Gordon, K. Eiglmeier, T. Garnier, F. Tekaia, E. Yeramian, S.T. Cole, Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex, in: G.F. Hatfull, W.R. Jacobs Jr (Eds.), Molecular Genetics of Mycobacteria, ASM Press, Washington, DC, 2000, pp. 19–36.

[4] D.B. Young, S.H.E. Kaufmann, P.W. Hermans, J.E. Thole, Mycobacterial protein antigens: a compilation, Mol. Microbiol. 6 (1992) 133–145.

[5] M.G. Sonnenberg, J.T. Belisle, Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry, Infect. Immun. 65 (1997) 4515–4524.

[6] G.G. Mahairas, P.J. Sabo, M.J. Hickey, D.C. Singh, C.K. Stover, Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*, J. Bacteriol. 178 (1996) 1274–1282.

[7] M.A. Behr, M.A. Wilson, W.P. Gill, H. Salamon, G.K. Schoolnik, S. Rane, P.M. Small, Comparative genomics of BCG vaccines by whole-genome DNA microarray, Science 284 (1999) 1520–1523.

[8] K.N. Lewis, R. Liao, K.M. Guinn, M.J. Hickey, S. Smith, M.A. Behr, D.R. Sherman, Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette–Guerin attenuation, J. Infect. Dis. 187 (2003) 117–123.

[9] A.S. Pym, P. Brodin, R. Brosch, M. Huerre, S.T. Cole, Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*, Mol. Microbiol. 46 (2002) 709–717.

[10] S. Ahmad, H.A. Amoudy, J.E.R. Thole, D.B. Young, A.S. Mustafa, Identification of a novel protein antigen encoded by a *Mycobacterium tuberculosis*-specific RD1 region gene, Scand. J. Immunol. 49 (1999) 515–522.

[11] M.J. Brennan, G. Delogu, The PE multigene family: a 'molecular mantra' for mycobacteria, Trends Microbiol. 10 (2002) 246–249.

[12] S.L. Sampson, P. Lukey, R.M. Warren, P.D. van Helden, M. Richardson, M.J. Everett, Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c, Tuberculosis (Edinb.) 81 (2001) 305–317.

[13] C. Espitia, J.P. Laclette, M. Mondragon-Palomino, A. Amador, J. Campuzano, A. Martens, M. Singh, R. Cicero, Y. Zhang, C. Moreno, The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? Microbiology 145 (1999) 3487–3495.

[14] S. Banu, N. Honore, B. Saint-Joanis, D. Philpott, M.C. Prevost, S.T. Cole, Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? Mol. Microbiol. 44 (2002) 9–19.

[15] D.C. Dillon, M.R. Alderson, C.H. Day, T. Bement, A. Campos-Neto, Y.A. Skeiky, T. Vedvick, R. Badaro, S.G. Reed, R. Houghton, Molecular and immunological characterization of *Mycobacterium tuberculosis* CFP-10, an immunodiagnostic antigen missing in *Mycobacterium bovis* BCG, J. Clin. Microbiol. 38 (2000) 3285–3290.

[16] R. Colangeli, J.S. Spencer, P. Bifani, A. Williams, K. Lyashchenko, M.A. Keen, P.J. Hill, J. Belisle, M.L. Gennaro, MTSA-10, the product of the Rv3874 gene of *Mycobacterium tuberculosis*, elicits tuberculosis-specific, delayed-type hypersensitivity in guinea pigs, Infect. Immun. 68 (2000) 990–993.

[17] F.X. Berthet, P.B. Rasmussen, I. Rosenkrands, P. Andersen, B. Gicquel, A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10), Microbiology 144 (1998) 3195–3203.

[18] P. Andersen, A.B. Andersen, A.L. Sorensen, S. Nagai, Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice, J. Immunol. 154 (1995) 3359–3372.

[19] A.L. Sorensen, S. Nagai, G. Houen, P. Andersen, A.B. Andersen, Purification and characterization of a low-molecular-mass T–cell antigen secreted by *Mycobacterium tuberculosis*, Infect. Immun. 63 (1995) 1710–1717.

[20] S.M. Arend, A. Geluk, K.E. van Meijgaarden, J.T. van Dissel, M. Theisen, P. Andersen, T.H. Ottenhoff, Antigenic equivalence of human T-cell responses to *Mycobacterium tuberculosis*-specific RD1-encoded protein antigens ESAT-6 and culture filtrate protein 10 and to mixtures of synthetic peptides, Infect. Immun. 68 (2000) 3314–3321.

[21] P.N. Brusasca, R. Colangeli, K.P. Lyashchenko, X. Zhao, M. Vogelstein, J.S. Spencer, D.N. McMurray, M.L. Gennaro, Immunological characterization of antigens encoded by the RD1 region of the *Mycobacterium tuberculosis* genome, Scand. J. Immunol. 54 (2001) 448–452.

[22] M. Braunstein, J.T. Belisle, Genetics of protein secretion, in: G.F. Hatfull, W.R. Jacobs Jr (Eds.), Molecular Genetics of Mycobacteria, ASM Press, Washington, DC, 2000, pp. 203–220.

[23] P.S. Renshaw, P. Panagiotidou, A. Whelan, S.V. Gordon, R.G. Hewinson, R.A. Williamson, M.D. Carr, Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence, J. Biol. Chem. 277 (2002) 21598–21603.

[24] F. Tekaia, S.V. Gordon, T. Garnier, R. Brosch, B.G. Barrell, S.T. Cole, Analysis of the proteome of *Mycobacterium tuberculosis* in silico, Tuber. Lung Dis. 79 (1999) 329–342.

[25] M.J. Pallen, The ESAT-6/WXG100 superfamily—and a new gram-positive secretion system? Trends Microbiol. 10 (2002) 209–212.

[26] D. Frishman, K. Albermann, J. Hani, K. Heumann, A. Metanomski, A. Zollner, H.W. Mewes, Functional and structural genomics using PEDANT, Bioinformatics 17 (2001) 44–57.

[27] S.B. Snapper, R.E. Melton, S. Mustafa, T. Kieser, W.R. Jacobs Jr, Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*, Mol. Microbiol. 4 (1990) 1911–1919.

[28] T. Parish, E. Mahenthiralingam, P. Draper, E.O. Davis, M.J. Colston, Regulation of the inducible acetamidase gene of *Mycobacterium smegmatis*, Microbiology 143 (1997) 2267–2276.

[29] E. Mahenthiralingam, P. Draper, E.O. Davis, M.J. Colston, Cloning and sequencing of the gene which encodes the highly inducible acetamidase of *Mycobacterium smegmatis*, J. Gen. Microbiol. 139 (1993) 575–583.

[30] S. Rozen, H.J. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, in: S. Krawetz, S. Misener (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa, NJ, 2000, pp. 365–386.

[31] T. Parish, J. Turner, N.G. Stoker, amiA is a negative regulator of acetamidase expression in *Mycobacterium smegmatis*, BMC. Microbiology 1 (2001) 19–23.

[32] J.A. Triccas, T. Parish, W.J. Britton, B. Gicquel, An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*, FEMS Microbiol. Lett. 167 (1998) 151–156.

[33] F. Romain, C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, G. Marchal, Deglycosylation of the 45/47-kDa antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses, Infect. Immun. 67 (1999) 5567–5572.

[34] C. Espitia, R. Mancilla, Identification, isolation and partial characterization of *Mycobacterium tuberculosis* glycoprotein antigens, Clin. Exp. Immunol. 77 (1989) 378–383.

[35] K.M. Dobos, K. Swiderek, K.H. Khoo, P.J. Brennan, J.T. Belisle, Evidence for glycosylation sites on the 45–kDa glycoprotein of *Mycobacterium tuberculosis*, Infect. Immun. 63 (1995) 2846–2853.

[36] T. Fifis, C. Costopoulos, A.J. Radford, A. Bacic, P.R. Wood, Purification and characterization of major antigens from a *Mycobacterium bovis* culture filtrate, Infect. Immun. 59 (1991) 800–807.

[37] T. Garbe, D. Harris, M. Vordermeier, R. Lathigra, J. Ivanyi, D. Young, Expression of the *Mycobacterium tuberculosis* 19–kDa antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation, Infect. Immun. 61 (1993) 260–267.

[38] W.R. Jacobs Jr, M. Tuckman, B.R. Bloom, Introduction of foreign DNA into mycobacteria using a shuttle plasmid, Nature 327 (1987) 532–535.

[39] J.E. Burlein, C.K. Stover, S. Offutt, M.S. Hanson, Expression of foreign genes in mycobacteria, in: B.R. Bloom (Ed.), Tuberculosis. Pathogenesis, Protection and Control, ASM Press, Washington, DC, 1994, pp. 239–252.

[40] C.K. Stover, V.F. de la Cruz, T.R. Fuerst, J.E. Burlein, L.A. Benson, L.T. Bennett, G.P. Bansal, J.F. Young, M.H. Lee, G.F. Hatfull, S.B. Snapper, R.G. Barletta, W.R. Jacobs Jr, B.R. Bloom, New use of BCG for recombinant vaccines, Nature 351 (1991) 456–460.

[41] Y.C. Manabe, J.M. Chen, C.G. Ko, P. Chen, W.R. Bishai, Conditional sigma factor expression, using the inducible acetamidase promoter, reveals that the *Mycobacterium tuberculosis* sigF gene modulates expression of the 16-kDa alpha-crystallin homologue, J. Bacteriol. 181 (1999) 7629–7633.

[42] M.E. Munk, S.M. Arend, I. Brock, T.H. Ottenhoff, P. Andersen, Use of ESAT-6 and CFP-10 antigens for diagnosis of extrapulmonary tuberculosis, J. Infect. Dis. 183 (2001) 175–176.

[43] J.A. Triccas, P.W. Roche, N. Winter, C.G. Feng, C.R. Butlin, W.J. Britton, A 35-kDa protein is a major target of the human immune response to *Mycobacterium leprae*, Infect. Immun. 64 (1996) 5171–5177.

[44] P.W. Roche, N. Winter, J.A. Triccas, C.G. Feng, W.J. Britton, Expression of *Mycobacterium tuberculosis* MPT64 in recombinant *M. smegmatis*: purification, immunogenicity and application to skin tests for tuberculosis, Clin. Exp. Immunol. 103 (1996) 226–232.

[45] M.S. Glickman, J.S. Cox, W.R. Jacobs, A novel mycolic acid cyclopropane synthetase is required for coding, persistence, and virulence of *Mycobacterium tuberculosis*, Mol. Cell 5 (2000) 717–727.